12 gennaio 2018 es.1 (Il parcheggio di Giulio)

Ogni sera Giulio cerca vicino a casa un parcheggio per la sua auto; la probabilità che lo trovi è 0.7; altrimenti parcheggia in una zona vietata, in questo caso rischia la multa, la cui probabilità è valutata in 0.2.

a) Qual è la probabilità che per 20 giorni consecutivi Giulio non prenda nessuna multa?

b) Questa mattina Giulio ha ripreso l'auto da dove l'aveva lasciata ieri; non ricorda se aveva parcheggiato regolarmente oppure in zona vietata, ma è sicuro di non avere preso la multa. Qual è la probabilità che il parcheggio di Giulio fosse in zona vietata?

Soluzione

a) In ciascuna giornata, l'evento "Giulio non prende la multa" si realizza se:
* Giulio trova un parcheggio in zona consentita (probabilità 0.7)
oppure
** Giulio parcheggia in zona vietata, ma evita la multa (probabilità $0.3 \cdot 0.8 = 0.24$).

Questi due eventi sono incompatibili; la probabilità della loro unione è $0.7 + 0.24 = 0.94$.

La probabilità che per 20 giorni consecutivi Giulio non prenda multe è

$$0.94^{20}$$

0.290106

b) Siano A, B gli eventi:

A) Giulio parcheggia in divieto di sosta
B) Giulio *non* prende la multa

Ci interessa $P(A \mid B)$.

$$P(A \mid B) = \frac{P(A \wedge B)}{P(B)} = \frac{P(A) \cdot P(B \mid A)}{P(B)} = \frac{0.3 \cdot 0.8}{0.94} \approx 0,255$$

12 gennaio 2018 es.2 (I pranzi di Dante)

Dante non pranza mai a casa sua, ma frequenta due locali, "Trattoria Paradiso" e "Osteria dell'Inferno"; la scelta tra uno e l'altro avviene in modo casuale, la decisione viene presa ogni giorno lanciando una moneta. Però, quando Dante va alla Trattoria Paradiso, vi è probabilità 0.4 che là incontri la sua amica Beatrice e sia invitato a casa di lei per il pranzo del giorno successivo; ogni volta che Beatrice ha Dante come ospite, con probabilità 0.6 lo inviterà a pranzo anche il giorno successivo; altrimenti Dante andrà in uno dei due locali abituali.

a) Descrivere attraverso una catena di Markov il luogo in cui Dante pranza in ciascun giorno (stato 1: a casa di Beatrice; 2: Trattoria Paradiso; 3: Osteria dell'Inferno). Classificare gli stati, stabilire se la catena è irriducibile, se è regolare e, in caso affermativo determinare la distribuzione invariante.

b) Un anno fa Dante ha pranzato all'Osteria dell'Inferno. Oggi, all'ora di pranzo, il suo amico Virgilio vorrebbe fargli visita; gli conviene cercarlo a casa di Beatrice, alla Trattoria Paradiso o all'Osteria dell'Inferno?

Soluzione

a) Il testo fornisce tutti i valori necesssari per compilare la matrice di transizione:

(da) \ (a)	1	2	3
1	0.6	0.2	0.2
2	0.4	0.3	0.3
3	0	0.5	0.5

; $p = \begin{pmatrix} 0.6 & 0.2 & 0.2 \\ 0.4 & 0.3 & 0.3 \\ 0 & 0.5 & 0.5 \end{pmatrix}$;

Tutti gli stati comunicano tra loro, e almeno un termine della diagonale principale è diverso da 0 (in effetti lo sono tutti e tre); quindi tutti gli stati sono ricorrenti e la catena di Markov è regolare. La distribuzione invariante si può trovare risolvendo il seguente sistema:

```
Solve[{{x1, x2, x3}.p == {x1, x2, x3}, x1 + x2 + x3 == 1}, {x1, x2, x3}]
```

$\{\{x_1 \to 0.333333, x_2 \to 0.333333, x_3 \to 0.333333\}\}$

dunque, la distribuzione invariante è quella uniforme. Questo risultato si poteva dedurre senza alcun calcolo, osservando che la matrice di transizione è *bistocastica* cioè, oltre alle righe, anche le sue colonne sono tali che la somma degli elementi è uguale a 1.

b) La distribuzione iniziale di probabilità (relativa a un anno fa) è $v_0 = (0, 0, 1)$, perché è noto che in quel momento lo stato era 3. Dopo 365 transizioni (un anno) la distribuzione di probabilità dei tre stati è $v_0 \cdot p^{365}$ (p è la matrice di transizione). Per il Teorema di Markov il vettore $v_0 \cdot p^n$ converge per $n \longrightarrow +\infty$ alla distribuzione stazionaria $\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$; per $n = 365$ il limite sarà in pratica già stato raggiunto, cosicché si può pensare $v_{365} = v_0 \cdot p^{365} \approx \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$; dunque Virgilio ha praticamente la stessa probabilità di trovare Dante in ciascuno dei tre posti in cui quest'ultimo consuma i suoi pasti.
La convergenza alla distribuzione stazionaria è in effetti molto rapida; già dopo una settimana la distribuzione di probabilità della catena è pressoché stabilizzata:

```
{0, 0, 1}.MatrixPower[p, 7]
```

```
{0.331968, 0.334016, 0.334016}
```

e ancora di più dopo due settimane:

```
{0, 0, 1}.MatrixPower[p, 14]
```

```
{0.333331, 0.333334, 0.333334}
```

12 gennaio 2018 es.3 (La previdenza di Ulisse)

Ulisse, di età 40, in partenza per un lungo viaggio, sottoscrive un'assicurazione sulla vita a favore di sua moglie Penelope (30): in caso di morte di Ulisse entro 10 anni da oggi, la Compagnia assicuratrice pagherà a Penelope l'importo di 50 000 € il giorno del suo primo compleanno che ella trascorrerà da vedova. Nulla sarà invece dovuto se Ulisse sopravvivrà per almeno 10 anni, oppure se nei 10 anni coperti dall'assicurazione Penelope dovesse morire prima di Ulisse oppure nello stesso anno.

a) Calcolare, esprimendola mediante i dati delle tavole di sopravvivenza, la probabilità che la Compagnia debba pagare a Penelope l'importo assicurato.

b) Calcolare, ancora in funzione dei dati di sopravvivenza e anche del tasso di mercato i, il premio puro che Ulisse deve pagare oggi per sottoscrivere l'assicurazione.

Si conviene per semplicità che Ulisse e Penelope compiano gli anni nello stesso giorno, e precisamente oggi; usiamo l'apice per gli indici demografici relativi alla popolazione femminile.

Soluzione

a) Calcoliamo la probabilità dell'evento contrario, ossia che l'assicurazione non sia tenuta ad alcun pagamento. Questa circostanza corrisponde all'unione dei due seguenti eventi incompatibili:

A) Ulisse è in vita tra 10 anni

B) Ulisse muore entro 10 anni, e Penelope muore nello stesso anno di Ulisse oppure prima.

$$P(A) = \frac{l_{50}}{l_{40}}.$$

L'evento B si verifica se Ulisse muore in età $40 + k - 1$, $1 \leq k \leq 10$, e Penelope non è in vita alla fine dell'anno in cui è deceduto Ulisse, cioè quando lei compirebbe $30 + k$ anni; la probabilità è quindi::

$$P(B) = \sum_{k=1}^{10} \frac{d_{40+k-1}}{l_{40}} \cdot \left(1 - \frac{l'_{30+k}}{l'_{30}}\right).$$

Infine, la probabilità che la probabilità che la Compagnia sia tenuta a qualche pagamento è

$$1 - (P(A) + P(B)) = 1 - \frac{l_{50}}{l_{40}} - \sum_{k=1}^{10} \frac{d_{40+k-1}}{l_{40}} \cdot \left(1 - \frac{l'_{30+k}}{l'_{30}}\right).$$

Altra possibile risoluzione, anche più semplice: l'evento "la Compagnia dovrà pagare" è l'unione di 10 eventi a due a due disgiunti, E_k $(1 \leq k \leq 10)$, con

$$E_k : \text{ Ulisse muore in età } 40 + x - 1 \ \textit{e} \text{ Penelope è vivente al compimento di } 30 + x \text{ anni}$$

Perciò la probabilità che la Compagnia sia tenuta al pagamento è

$$\sum_{k=1}^{10} P(E_k) = \sum_{k=1}^{10} \frac{d_{40+k-1}}{l_{40}} \cdot \frac{l'_{30+k}}{l'_{30}} .$$

b) Per $1 \le k \le 10$, allo scadere di k anni da oggi l'Assicurazione dovrà pagare un importo aleatorio X_k il cui valore è di 50 000 €, se Ulisse è deceduto in età $40 + k - 1$ e Penelope è in vita al compimento dell'età $30 + k$, oppure 0 €, altrimenti. Allora:

$$E(X_k) = 50\,000 \cdot \frac{d_{40+k}}{l_{40}} \cdot \frac{l'_{30+k}}{l'_{30}}$$

Il premio puro che Ulisse deve pagare oggi per sottoscrivere l'assicurazione è la somma dei valori attuali delle $E(X_k)$, cioè

$$50\,000 \cdot \sum_{k=1}^{10} \frac{d_{40+k}}{l_{40}} \cdot \frac{l'_{30+k}}{l'_{30}} \cdot (1+i)^{-k}.$$

12 gennaio 2018 es.4 (I pistacchi di Dimitri)

Il seme commestibile dei pistacchi è contenuto in un guscio legnoso costituito di due parti (similmente a una noce); queste possono essere separate da una fenditura, che facilita l'apertura ("guscio aperto"), oppure no ("guscio chiuso"); in questo caso l'apertura necessita di uno strumento, per esempio uno schiaccianoci. I pistacchi sono considerati più pregiati quanto più piccola è la percentuale di quelli "chiusi". Quando un produttore va al mercato, l'acquirente verifica il prodotto prendendone una manciata, sulla quale controlla la proporzione di "chiusi"; su questa base viene perfezionata la vendita e stabilito il prezzo.

a) Dimitri coltiva pistacchi; al mercato il suo acquirente ne prende 150 e osserva che 21 di essi sono chiusi. Determinare un intervallo di confidenza al 90%, centrato nel valore campionario, per la probabilità p che ciascuno dei pistacchi del raccolto di Dimitri sia chiuso.

b) Con i risultati sperimentali di (a), cioè 21 pistacchi chiusi su 150, qual è il livello di confidenza dell'intervallo [0.12 , 0.16], per il valore di p?

Soluzione

a) Lo stimatore di p è in questo caso $\overline{x} = \frac{21}{150} = 0.14$

$$\overline{\mathbf{x}} = \frac{\mathbf{21.}}{\mathbf{150}}$$

```
0.14
```

L' intervallo di confidenza al livello $1 - \alpha = 0.90$ per p è $\left[\overline{x} - \phi_{1-\frac{\alpha}{2}} \sqrt{\frac{\overline{x}(1-\overline{x})}{150}}, \overline{x} + \phi_{1-\frac{\alpha}{2}} \sqrt{\frac{\overline{x}(1-\overline{x})}{150}}\right]$ con $\alpha = 0.1$ e $\phi_{1-\frac{\alpha}{2}}$ quantile di livello $1 - \frac{\alpha}{2}$ per la distribuzione normale standard $N(0, 1)$. Attualmente $\alpha = 0, 1$ quindi ci serve

```
ϕ0.95 = Quantile[NormalDistribution[0, 1], 0.95]
```

```
1.64485
```

Gli estremi dell'intervallo di confidenza cercato sono:

$$\left\{\overline{\mathbf{x}} - \phi_{0.95} \sqrt{\frac{\overline{\mathbf{x}}\,(1-\overline{\mathbf{x}})}{150}}, \overline{\mathbf{x}} + \phi_{0.95} \sqrt{\frac{\overline{\mathbf{x}}\,(1-\overline{\mathbf{x}})}{150}}\right\}$$

```
{0.093399, 0.186601}
```

b) L'intervallo, centrato rispetto allo stimatore $\overline{x}$, va pensato ancora nella forma

$\left[\overline{x} - \phi_{1-\frac{\alpha}{2}} \sqrt{\frac{\overline{x}(1-\overline{x})}{150}}, \overline{x} + \phi_{1-\frac{\alpha}{2}} \sqrt{\frac{\overline{x}(1-\overline{x})}{150}}\right]$ in cui ora è incognito α ; cerchiamo quindi il quantile ϕ per $N(0, 1)$ tale che sia $\phi \cdot \sqrt{\frac{\overline{x}(1-\overline{x})}{150}} = 0.02$

```
Solve[ϕ * Sqrt[(x̄ (1 - x̄))/150] == 0.02, ϕ]
```

```
{{ϕ → 0.705931}}
```

Ciò significa che $1-\frac{\alpha}{2}=\Phi(0.705931)$ dove Φ rappresenta la funzione ripartizione di $N(0, 1)$; quindi il valore di $1-\frac{\alpha}{2}$ è

```
CDF[NormalDistribution[0, 1], 0.7059312080025175`]
```

```
0.759885
```

e ora calcoliamo α:

```
Solve[1 - α/2 == %, α]
```

```
{{α → 0.480231}}
```

Il livello di confidenza dell'intervallo di (b) è $1-\alpha$, cioè

```
1 - 0.48023091880553626`
```

```
0.519769
```

ossia circa 52%; non è un livello molto buono

12 gennaio 2018 es.5 (Ancora pistacchi)

Dimitri (vedi es.4) è convinto che non più del 10% dei suoi pistacchi siano chiusi, cioè che la probabilità di ciascun pistacchio di essere "chiuso" sia $p \leq 0.10$. Il suo cliente concluderà l'acquisto soltanto se un test unilaterale su 150 pistacchi al livello del 5% consentirà di respingere l'ipotesi H: «$p \geq 0.15$».

a) Calcolare la probabilità che l'ipotesi H sia respinta, se è veritiera la stima che Dimitri attribuisce a *p.*

b) Quale migliore valutazione di *p* dovrebbe avere Dimitri affinché la probabilità di superare il test sia almeno del 90%? (È sufficiente scrivere un'equazione risolvendo la quale si otterrebbe il massimo valore di *p* soddisfacente il requisito, tralasciando il calcolo risolutivo).

Soluzione

a) Una statistica-test per il problema considerato è il numero *k* di pistacchi chiusi che saranno trovati tra i 150 del campione. La regione di rifiuto per l'ipotesi H al livello $\alpha = 0.05$ è $[0, k_0]$ con k_0 tale che

$$\frac{k_0+\frac{1}{2}-150\cdot 0.15}{\sqrt{150\cdot 0.15\cdot 0.85}} = -\phi_{0.95}$$

con $\phi_{0.95}$ quantile di livello 0.95 per $N(0, 1)$:

```
ϕ0.95 = Quantile[NormalDistribution[0, 1], 0.95]
```

```
1.64485
```

Ricaviamo k_0:

```
Solve[(k0 + 1/2 - 150 * 0.15)/Sqrt[150 * 0.15 * 0.85] == -ϕ0.95, k0]
```

```
{{k0 → 14.8067}}
```

e quindi, tenendo conto che *k* assume valori interi, la regione di rigetto dell'ipotesi H è $D = [0, 14]$.

La probabilità che l'ipotesi H sia respinta, se si suppone che sia $p \leq 0.10$, è maggiore o uguale della *potenza* del test relativa alla regione D calcolata per $p = 0.10$, cioè di $\pi_D(0.10) = P^{0.10}(k \leq 14)$, che vale

```
CDF[NormalDistribution[0, 1], (14.5 - 150 * 0.10)/Sqrt[150 * 0.10 * 0.90]]
```

```
0.445878
```

b) Se la probabilità di ciascun pistacchio di essere chiuso è p, allora la variabile X che conta quanti pistacchi sono chiusi in un campione di 150 ha distribuzione binomiale $B(150, p)$. L'evento desiderato è $X \leq 14$, equivalente a $X \leq 14.5$ perché X assume valori interi. $Z = \frac{X-150\,p}{\sqrt{150\,p(1-p)}}$ è approssimativamente distribuita come $N(0, 1)$ e si ha

$$P(X \leq 14.5) = P\left(Z \leq \frac{14.5-150\,p}{\sqrt{150\,p(1-p)}}\right) = \Phi\left(\frac{14.5-150\,p}{\sqrt{150\,p(1-p)}}\right)$$

e questa espressione è funzione decrescente di p. Se desideriamo che questa probabilità sia ≥ 0.90, bisogna che $\frac{14.5-150\,p}{\sqrt{150\,p(1-p)}} \geq \phi_{0.90}$ (quantile di livello 0.90 per $N(0, 1)$):

```
In[5]:= ϕ0.90 = Quantile[NormalDistribution[0, 1], 0.90]
Out[5]= 1.28155
```

quindi bisogna che sia $p \leq \vartheta$ con ϑ tale che

$$\frac{14.5-150*\vartheta}{\sqrt{150*\vartheta*(1-\vartheta)}} = \phi_{0.90}$$

Questa equazione, se risolta, fornisce il valore richiesto per ϑ. Esso è

```
In[6]:= Solve[(14.5 - 150 * ϑ)/Sqrt[150 * ϑ * (1 - ϑ)] == ϕ0.90, ϑ]
Out[6]= {{ϑ → 0.0699732}}
```

vale a dire che Dimitri potrà affrontare il test con probabilità almeno 90% di superarlo, se sarà sicuro che la percentuale di pistacchi chiusi del suo raccolto non superi il 7%.

12 gennaio 2018 es.6 (Una regione trilingue)

In una regione di confine di un determinato Paese ci sono tre gruppi linguistici, di madrelingua “x”, “y”, “z” rispettivamente. Si vuole stabilire se c'è oppure no correlazione tra il gruppo linguistico di appartenenza e la collocazione lavorativa dei cittadini; con questo scopo viene selezionato un campione di $n = 400$ cittadini in età lavorativa, in cui sono rappresentati i tre gruppi linguistici, e viene classificato il loro impiego nelle categorie “dirigente”, “impiegato”, “operaio”; i risultati ottenuti sono qui riassunti.

10.	60.	30.	madrelingua x
20.	50.	50.	madrelingua y
10.	90.	80.	madrelingua z
dirigenti	impiegati	operai	□

a) Stabilire se, al livello 1%, si può credere che l'attività lavorativa dei cittadini di quella Regione sia indipendente dal gruppo linguistico di appartenenza.

b) Stabilire mediante un test bilaterale se, al livello del 5%, è da rifiutare oppure no l'ipotesi

H : le probabilità p_x, p_y di ottenere un lavoro da dirigente per i cittadini di madrelingua x oppure y sono uguali.

Soluzione

La seguente matrice aggiunge ai dati del problema le distribuzioni marginali relative a "lingua" e "lavoro":

$$\begin{pmatrix} \square & \text{dirigenti} & \text{impiegati} & \text{operai} & \text{marg.lingua} \\ \text{madrelingua x} & 10. & 60. & 30. & 100. \\ \text{madrelingua y} & 20. & 50. & 50. & 120. \\ \text{madrelingua z} & 10. & 90. & 80. & 180. \\ \text{marg.lavoro} & 40. & 200. & 160. & 400. \end{pmatrix}$$

Questa corrisponde alle seguenti frequenze relative (accoppiate e marginali)

```
Print[MatrixForm[m1f]]
```

$$\begin{pmatrix} 0.025 & 0.15 & 0.075 & 0.25 \\ 0.05 & 0.125 & 0.125 & 0.3 \\ 0.025 & 0.225 & 0.2 & 0.45 \\ 0.1 & 0.5 & 0.4 & 1. \end{pmatrix}$$

Invece la matrice con le frequenze relative accoppiate teoriche in caso di indipendenza, cioè i prodotti delle distribuzioni marginali è

```
Print[MatrixForm[m2f]]
```

$$\begin{pmatrix} 0.025 & 0.125 & 0.1 & 0.25 \\ 0.03 & 0.15 & 0.12 & 0.3 \\ 0.045 & 0.225 & 0.18 & 0.45 \\ 0.1 & 0.5 & 0.4 & 1. \end{pmatrix}$$

La statistica-test per l'indipendenza delle due variabili "secondo" e "contorno" è

$$\mathbf{t = n * \sum_{h=1}^{3} \sum_{k=1}^{3} \frac{(m1f[[h]][[k]] - m2f[[h]][[k]])^2}{m2f[[h]][[k]]}}$$

16.0278

Se le due variabili considerate sono indipendenti, t ha aprossimativamente distribuzione $\chi^2(4)$ (i gradi di libertà sono (3-1)×(3-1)). Il quantile che limita inferiormente la regione di rifiuto del'ipotesi di indipendenza è

```
q = Quantile[ChiSquareDistribution[4], .99]
```

13.2767

L'ipotesi viene rifiutata.

b) La proporzione dei dirigenti tra i madrelingua *x* nel nostro campione è

$$\bar{x} = \frac{10}{100.}$$

0.1

mentre per i madrelingua *y* è

$$\bar{y} = \frac{20}{120.}$$

0.166667

La statistica-test per questo problema è

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\bar{x}(1-\bar{x})}{100} + \frac{\bar{y}(1-\bar{y})}{120}}}$$

-1.46977

Se l'ipotesi H è vera allora t ha approssimativamente distribuzione $N(0, 1)$ e allora la regione di rigetto dell'ipotesi H al livello $\alpha = 0.05$ è $D =]-\infty, -\phi_{1-\frac{\alpha}{2}}[\cup]\phi_{1-\frac{\alpha}{2}}, +\infty[$ dove è

$\phi_{1-\frac{\alpha}{2}}$ **= Quantile[NormalDistribution[0, 1], 0.975]**

1.95996

Gli attuali valori sperimentali non consentono di respingere l'ipotesi H, al livello del 5%.